1: *Microb Comp Genomics* 1996;1(3):129-50

Phylogenetic approaches to the identification and characterization of protein families and superfamilies.

Saier MH

Department of Biology, University of California at San Diego, La Jolla, USA.

With the advent of megabase genome sequencing, the need for computational analyses increases exponentially. Sequencing errors must be corrected, encoded proteins must be identified, functions must be assigned to these proteins, and distant phylogenetic relationships must be recognized in order to maximize the yield of information obtainable from genome sequencing projects. Both the computer and the human brain have their limitations, but using them in combination, the biologist can vastly extend his or her analytic capabilities. Computer techniques can be used to estimate protein structure, function, biogenesis, and evolution. In this review, the application of available computer programs to several protein families, particularly transport, receptor, and transcriptional regulatory protein families, illustrate our current capabilities and limitations. Although some multidomain protein families are evolutionarily homogeneous, others have mosaic origins. Evidence concerning the nature and frequency of occurrence of domain shuffling, splicing, fusion, deletion, and duplication during evolution of specific protein families is evaluated. It is shown that specific families of enzymes, receptors, transport proteins, and transcriptional regulatory proteins share a common evolutionary origin, frequently diverging in function because of domain splicing and ligation. Some large families arose gradually over evolutionary time, whereas others developed suddenly, due to bursts of intragenic or intergenic (or both) duplication events occurring over relatively short periods of time. It is argued that energy coupling to transport was a late occurrence, superimposed on preexisting mechanisms of solute facilitation. It is also shown that several transport protein families have evolved independently of each other, employing different routes, at different times in evolutionary history, to give topologically similar transmembrane protein complexes.